Dynamic User Profiling Model Construction for Social Academic Apps Based on Small Data (Postprint)

Authors: Zhang Liman, Zhang Xiangxian, Wu Yawei, Guo Shunli

Date: 2023-04-01T16:15:50+00:00

Abstract

[Purpose/Significance] Constructing a dynamic user profiling model for social academic Apps based on small data provides ideas and references for such platforms to effectively predict user behavior evolution trends and improve precision service levels. [Methods/Process] First, based on an in-depth analysis of the concept and characteristics of small data and combined with the features of social academic Apps, a dynamic profiling tag system is designed from two aspects: user surface behavior and deep-seated driving factors; second, small data that is strongly correlated with users and of high value is collected as data support for profiling, and the acquisition and processing methods of profiling small data are clarified; finally, research methods for implementing dynamic profiling are proposed and an overall framework model is formed. [Results/Conclusion] Constructing dynamic user profiles for social academic Apps based on small data can effectively refine profiling granularity, overcome the lagging drawbacks of previous profiling methods, and holds important reference value for social academic App platforms to enhance precision service levels in data-driven contexts.

Full Text

Preamble

Volume 64, Issue 5, March 2020
ChinaXiv Cooperative Journal

Research on Constructing a Dynamic User Portrait Model for Social Academic App Users Based on Small Data

Zhang Liman[1], Zhang Xiangxian[1], Wu Yawei[1], Guo Shunli[2]
[1] School of Management, Jilin University, Changchun 130022
[2] School of Media, Qufu Normal University, Rizhao 276826

Abstract:
[Objective/Significance] This study constructs a dynamic user portrait model for social academic App users based on small data, providing ideas and references for social academic App platforms to effectively predict the evolution trend of user behavior and improve precision service levels. [Method/Process] First, based on an in-depth analysis of the concept and characteristics of small data, combined with the features of social academic Apps, a dynamic portrait label system is designed from two aspects: user surface behavior and deep driving factors. Second, small data strongly correlated with and of high value to users is collected as the data support for the portrait, and the acquisition and processing methods of portrait small data are clarified. Finally, the research method for implementing dynamic portraits is proposed and an overall framework model is formed. [Result/Conclusion] Constructing dynamic portraits of social academic App users based on small data can effectively refine portrait granularity and overcome the lag shortcomings of previous portraits, which has important reference value for improving the precision service level of social academic App platforms in data-driven contexts.

Keywords: small data; social academic App; dynamic user portrait; behavior prediction
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2020.05.006

Introduction

With the development of the Internet and rapid information updates, traditional offline communication models can no longer meet scholars' diversified and personalized knowledge needs or their professional and timely service requirements. Social academic mobile applications (hereinafter referred to as "social academic Apps") represented by DXY and Xiaomuchong have become new avenues for research users to acquire knowledge resources and conduct academic exchanges [1]. As emerging platforms that rely on user interaction behaviors, sustained user engagement is key to their operation and development. Therefore, how to identify the dynamic evolution of user behavior and provide precise services according to different user characteristics has become a major challenge for platform operations.

User portraits, as effective tools for delineating target users and improving decision-making efficiency, have been widely applied in many fields. R. J. Holden et al. constructed health personas for elderly users from dimensions such as age, gender, and economic background [2]; M. Trusov et al. depicted consumer interest preferences by analyzing profile and behavioral data [3]. Such portraits based on user basic attributes and behavioral characteristics provide guidance for identifying typical groups, but the coarse-grained "one-size-fits-all" role models cannot uncover users' deep needs. Therefore, some scholars have attempted to build more comprehensive and accurate portrait models from the perspective of individual user small data: Chen Chen et al. constructed a precise portrait model for library users oriented toward personalized services by collecting library user small data [4]; Sun Danxia et al. argued that relying on small data can create vivid and comprehensive "user self-portraits" based on comprehensive behavioral characteristics and context awareness [5]. Introducing small data that comprehensively represents individual user characteristics can effectively refine portrait granularity. However, current portraits built on small data mostly collect data at a certain time node to create instantaneous portraits, where the generated role model is in a relatively static state that only displays users' current behavioral characteristics and cannot make reasonable inferences about their future behavioral trends. Moreover, facing data surges, instantaneous static portraits often adopt repetitive iteration methods to re-depict the full user picture, which is inefficient, fails to fully utilize previous portrait information, and has limited practical value. The few exploratory ideas for user dynamic portraits mostly focus on the library field [6-7], and there is still a lack of application attempts in the context of social academic Apps.

In summary, constructing fine-grained user dynamic portraits is an important way to fully present the complete picture of users, gain real-time insights into user behavior evolution, and improve platform precision service levels. Therefore, this study takes social academic Apps as the research object and constructs a dynamic portrait model that breaks through surface behavioral differences based on small data that is strongly correlated with social academic App users, of high value, and comprehensive. On the basis of describing user behavioral characteristics, it explores the driving factors and triggers behind their behavior, outlines user roles with stability, continuity, and dynamism, enabling platform operators to deeply understand user behavioral needs and evolution trends, and thus proactively propose personalized operation strategies to facilitate precise matching between products and users.

---

1 Related Research and Problem Statement

1.1 Small Data

The concept of small data was initially discovered and proposed by Professor D. Estrin of Cornell University, who believed that comprehensive tracking of users' daily behavioral data could dynamically monitor their health changes [8]. Currently, there is no clear and unified definition of small data in academia, but it is generally recognized that small data is the complete data collection of all-round, multi-level behavioral patterns and context awareness centered on individuals or teams [9]. Over time, these data collections continue to enrich, providing strong support for dynamically mining user needs, preferences, and behavioral patterns. Current research on small data focuses on personalized recommendation [10], precision services [6], interest discovery and prediction

[11], and theoretical discussions on small data fusion [12], but there is still a lack of conceptual frameworks for constructing academic App portrait models based on small data.

Through reviewing research and applications related to small data, this study identifies the following typical characteristics: **User-centricity**. Unlike big data that focuses on macro aggregates, small data is individualized data centered on users that can reveal their true inner selves, with higher value density, providing support for precisely describing users' comprehensive behavioral characteristics and complete conceptual pictures; **Multidimensional complexity**. Compared with big data, small data emphasizes uninterrupted, multi-dimensional, deep collection of individual panoramic data and correlation with contextual factors, with broader data sources and more diverse data types, thus requiring the integration of multiple data processing methods; **Focus on causal relationships**. Big data focuses on surface-level correlation descriptions between data without exploring the deep reasons affecting data correlations, whereas small data not only focuses on data correlation manifestations but also emphasizes revealing the driving factors behind data relationships. In addition, small data also shares characteristics of big data such as value, dynamism, and rapidity, and can be regarded as a supplement and extension of big data, thus allowing full borrowing of big data-related technologies for small data processing and utilization.

1.2 Dynamic Portrait of Social Academic App Users

As integrated platforms that balance academic and social functions, social academic Apps refer to applications installed on mobile smart terminals that provide academic resources or communication platforms for users [13]. Currently, due to the popularity of mobile smart terminals and widespread use of Apps, profiling App users has attracted attention from scholars in multiple fields. For example, Ji Qingnan obtained pain points of smart bus App users through establishing user portraits and emotion fluctuation charts [14]; Li Dawei et al. designed a personalized recommendation model for book recommendation Apps based on user portraits and collaborative filtering algorithms [15]; Han Zhangjunjie constructed user portraits for information Apps as an entry point, using clustering algorithms and association rules to divide user groups and mine group characteristics, aiming to optimize precision service models [16]. The above research provides references for scholars and operators to locate group needs, achieve personalized recommendations, and implement precision marketing through user portraits. However, social academic Apps are still relatively new, and relevant research results are scarce. Previous studies mainly explored their technical development methods [17], user usage [18], or adoption influencing factors [13] through theoretical discussions or questionnaire surveys, lacking data-driven exploration attempts to segment user groups and grasp user demand evolution through user portraits.

Traditional user portraits only collect data at a certain time node, abstracting a model that statically displays users' current and historical full picture based on behavioral characteristics and lifestyle habits [16]. The dynamic portrait of social academic App users refers to introducing time segments on the basis of depicting users' complete conceptual picture, using scientific methods to dynamically and continuously outline the trajectory of user behavior during interaction with the platform. From this perspective, static portraits can be seen as frozen descriptions of dynamic portraits at a certain time node. Some scholars have made active attempts at dynamic portrait concepts: Liu Yong et al. predicted user interest change trends based on historical interaction data for dynamic recommendation [19], but ignored users' subjective variability, and recommendations based on historical data may not necessarily meet users' current or future needs; Wang Yicheng et al. believed that user behavior big data could be collected to build a behavior tag library, and then the portrait model could be continuously corrected based on user feedback [20], but still only focused on surface behavior without considering the deep motivation behind behavior, resulting in coarse portrait granularity. Zhang Huimin discussed the necessity and constituent dimensions of dynamic user portraits in the context of lifestyle transformation, but focused on analyzing the requirements for interaction designers in building dynamic portraits [21], while not considering the evolution law of portraits over time. Therefore, existing user dynamic portrait methods suffer from coarse granularity and poor timeliness, and research on user dynamic portraits tailored to the characteristics of social academic Apps is very scarce.

1.3 Characteristics of Social Academic App User Dynamic Portrait Based on Small Data

Social academic App users are mostly highly educated researchers with professional research fields. Compared with general Apps, their demand characteristics are more obvious, and the convenience of mobile terminals helps meet users' instant and contextual academic and social needs. Therefore, when designing the portrait tag system, user behavioral characteristics and deep driving factors should be comprehensively, multi-dimensionally, and deeply mined, while considering users' states and needs in specific contexts. In addition, dynamic portraits require data continuity that can continuously and stably reveal users' characteristic trends. According to the concept of small data, small data for social academic App users refers to various data collections centered on the entire process of users using the platform, which can truly and comprehensively represent users' fine-grained characteristics such as behavior, motivation, and usage context. Obtained through long-term monitoring of individual user behavioral characteristics, it can meet the data continuity requirements for building dynamic portraits.

Compared with building user dynamic portraits based on big data, social academic App user dynamic portraits based on small data have the following typical characteristics: **Precision**. Big data mainly originates from behavioral activities of large numbers of users, with portraits focusing on the same type of characteristics of massive users. Small data originates from various activities of individual users, focusing on multi-dimensional characteristics of single users. Portrait models based on small data can more precisely display users' complete conceptual picture; **Deep-level**. Common big data portraits are mostly constructed by collecting users' basic attributes and behavioral data, suitable for research scenarios that conduct preliminary division of user groups. Small data portraits also emphasize highlighting behavioral driving factors, which is more conducive to deep mining of user dynamic behavior patterns and timely formulation of personalized operation strategies; **Practicality**. Big data has huge volume and low value density, containing large amounts of invalid data that interfere with portrait accuracy, resulting in low efficiency of big data-based dynamic portraits. Small data focuses on targeted individual users with moderate data volume. Moreover, the small data collection process is conducted in a relatively closed environment, making it easier to establish good communication mechanisms with users, reduce concerns about privacy leakage, and obtain high-value data. Therefore, dynamic portrait models based on small data have higher practical value.

In summary, dynamic portraits based on small data fully align with the strategic goals of social academic App platforms to efficiently track user characteristic trends and timely formulate precision marketing strategies. Therefore, this study explores the construction of a dynamic portrait model for social academic App users from the perspective of small data. First, it clarifies the overall process of building dynamic portraits for social academic App users based on small data, then uses Lewin's field theory as the basis to determine the dimensions of dynamic portraits, and accordingly proposes the collection and processing methods of relevant small data. Next, it explores suitable methods for outlining dynamic portraits, and finally forms an overall framework model for social academic App user dynamic portraits based on small data, providing new research perspectives and reference ideas for precision operation of social academic App platforms in data-driven contexts.

---

2 Overall Design of Social Academic App User Dynamic Portrait Based on Small Data

Typical user portrait construction methods include A. Cooper's "Seven-Step Persona Method" and L. Nielsen's "Ten-Step Persona Method," which can be summarized into three stages: obtaining tag data, segmenting user groups, and establishing and enriching user portraits [22]. According to the concept of dynamic portraits for social academic App users, time segments must be introduced in the construction process to discover the dynamic trajectory of user behavior based on the migration relationships between adjacent time period clusters. Moreover, the dynamic portrait model proposed in this study emphasizes the pertinence and three-dimensionality of the tag system. Therefore, the process of building social academic App user dynamic portraits based on small data should be divided into four stages, as shown in [Figure 1: see original paper]:

Stage 1: From the perspective of small data, combined with the characteristics of social academic Apps, design a three-dimensional dimension tag system that includes deep driving factors of user behavior.

Stage 2: According to the dimension tags, collect user small data and perform preprocessing.

Stage 3: Divide data according to time information in user small data, use clustering algorithms to cluster data within specific time segments, divide users into different groups, construct time-specific portraits, and store them in a database.

Stage 4: Based on time-specific portraits, determine the cluster centers in each time period (cluster centers can be regarded as typical representatives of user groups in that period). Mine and analyze the dynamic migration relationships of cluster centers in adjacent time periods to track user behavior change trajectories and achieve the depiction of user dynamic portraits.

The above design process not only references the basic stages of traditional user portraits but also integrates the particularity and requirements of dynamic portraits based on small data proposed in this study, ensuring practical value and innovative effect.

---

3 Tag System and Small Data Acquisition for Social Academic App User Dynamic Portrait

3.1 Theoretical Basis for Tag Dimension Determination

User psychological activities are the internal factors supporting behavior generation, and any behavior is driven by certain intentions before occurrence [23]. Social psychologist Lewin proposed field theory to analyze the driving forces behind individual behavior generation and behavior change processes. Field theory includes field theory and dynamics theory. Field theory defines "field" as the overall form of interdependence between individuals and the environment, also known as individual life space (LS). Individuals' psychology and behavior always occur and move within this space, which can be expressed by the function formula:

$$B = f(P * E) = f(LS) \quad (1)$$

where B represents externalized behavior, P represents individual internal needs, E represents the psychological environment (i.e., scenarios that stimulate internal needs), and f is the function of interaction between individuals and the environment [24]. Therefore, field theory believes that individual behavior is the result of interaction between subjects and contexts. Dynamics theory proposes that the driving force of individual psychology or behavior stems from the tension generated during the interaction between individuals and contexts. That is, when individual needs are not met, their psychology will be in a state of tension, driving behavior generation to alleviate or eliminate psychological tension. In addition, individuals' psychological goals are also important factors driving behavior generation.

According to field theory, social academic App user behavior is driven by both internal needs and external contexts. Specifically, users often have an intuitive and basic internal need for social academic Apps, such as the need to find literature or conduct research cooperation. The scenarios users are in or interaction scenarios with other users can also generate needs. When users find these needs cannot be met, they will experience psychological tension, leading to a series of behavioral activities attempting to eliminate the psychological tension, such as searching, downloading needed knowledge, or asking questions and collecting content to express personal demands and interest preferences. Therefore, social academic App user behavior, internal needs, and external contexts present a dynamic interactive relationship, providing a theoretical basis for this study to deeply analyze the user behavior driving process and determine dynamic portrait dimensions.

3.2 Composition of Tag System and Small Data Acquisition

Based on field theory and the characteristics of social academic Apps, this study believes that factors driving social academic App user behavior include user value orientation, cognitive ability, contextual features, and social relationships. Combined with two basic portrait factors—user natural attributes and behavior preferences—a six-dimensional portrait system is constructed, as shown in [Figure 2: see original paper]. The internal relationships among the dimensions are:

Natural attributes and behavior preferences form the basic framework for outlining portraits. Value orientation, cognitive ability, contextual features, and social relationships drive behavior generation, where value orientation and cognitive ability belong to user self-driving (internal driving force P), while context and social interaction are external situational stimulus factors (inducement E). Behavior preference is the externalized manifestation of the combined effects of natural attributes, value orientation, cognitive ability, contextual features, and social relationships.

(1) Natural Attributes. Natural attributes with persistent and stable characteristics are the foundation for generating user behavior [25]. Gender and age are influencing factors of group behavior, preferences, and demand trends. Since social academic Apps aim to provide knowledge and communication, users can obtain corresponding content and services according to their interest needs, so users' education level and professional fields should also be considered.

(2) Behavior Preferences. User behavior preferences refer to users' tendencies toward certain things and their degree of attention, which is the externalization of needs [22]. Grasping user behavior preferences requires consideration of both intuitive demands and interest tendencies. Users often acquire needed knowledge through searching, browsing, clicking, and downloading, or intuitively express their knowledge demands through posting and feedback. Following interesting users and collecting valuable content reflect interest tendencies.

(3) Value Orientation. Self-discrepancy theory posits that the ideal self is the level of capability characteristics an individual expects to possess, representing a vision for one's future state, while the actual self refers to capability characteristics currently possessed. The gap between them drives users to continuously generate behaviors to narrow the gap [26]. Research users use social academic Apps to improve their capabilities through solving research problems or mutual research assistance, constantly approaching their ideal selves. Therefore, value orientation is the source of power for user behavior generation, which can be examined from both user vision and self-assessment.

(4) Cognitive Ability. Cognitive ability refers to users' ability to identify, process, and effectively apply information content. Different cognitive abilities drive users to generate differentiated behaviors [27]. Cognitive ability is related to both user quality and cultural level and platform interaction degree. Users' cognitive abilities gradually improve through interaction with the platform, and their value contribution to the platform increases. Therefore, cognitive ability can be considered from both users' own situations and their contribution value to the platform. Among them, level and certified identity reflect users' own levels, while the number of followers, collections, and likes reflect their contribution value to the platform.

(5) Contextual Features. Changes in user needs under different contexts lead to changes in behavior preferences [28]. Context is dynamic and continuous, and the linking of contextual fragments forms the life trajectory of research users. Therefore, analyzing contextual factors is an inevitable requirement for building developmental user dynamic portraits and timely responding to user needs. According to context classification [29] and research object characteristics, this study believes that contextual features can be perceived from four aspects: time context, location context, user context, and device context.

(6) Social Relationships. While meeting users' knowledge needs, social academic Apps also have social functions, encouraging users to actively participate in knowledge exchange, sharing, and innovation [25]. Users on the platform can further explore their potential needs and interests through following, discussing, and sharing with others, thereby affecting their App usage behavior patterns. Therefore, social relationships can be regarded as a group network context that drives user behavior patterns. The groups users join, the number of people they follow, and interaction data are important tags for dynamically and deeply tracking user behavior trajectories.

Users' basic information and social relationship data are stored in the App management backend. Behavior preferences can be obtained from behavior data and user-generated content stored in user logs. Value orientation can be mined from user-generated content, questionnaires, or interviews. Cognitive ability can be obtained from certification materials submitted by users and interaction data analysis. Perception of contextual features mainly relies on sensors, positioning systems, and smart wearable devices. Social relationships are obtained through log mining and social network analysis. The tag system and small data acquisition methods for social academic App user dynamic portraits based on small data are shown in .

3.3 Small Data Processing

Previous user portrait data mostly consisted of basic attribute data and behavioral data, which could be transformed through coding or simple processing for direct experimental analysis [3]. However, the portrait constructed in this study also needs to integrate review or post text (i.e., content features) on this basis. Literature [30] proposed a behavior-content fusion model for portraits, which first concatenates user-posted text, then performs deep user representation learning, and finally obtains category tags through clustering as a feature to be added to behavioral features for portrait depiction. This method provides a reference for multi-source data processing but does not consider the topic features of text content. The professionalism and domain specificity of social academic Apps are prominent, and users' posts, replies, or followed topics are often related to their specific fields. This characteristic makes it possible to mine and analyze users' interest fields from text content. Therefore, this study proposes a text data modeling method based on the LDA topic model, as shown in [Figure 3: see original paper].

Step 1: Collect text content, concatenate and store it as text documents by user ID. After cleaning, import a domain lexicon for Chinese word segmentation and stop word filtering, cutting the original text into sequences of feature words. Then use TF-IDF to calculate feature word weights from both word frequency and importance perspectives, retaining important feature words.

Step 2: Use the Gibbs sampling in the LDA (Latent Dirichlet Allocation) topic model to mine hidden topics in the text. That is, according to statistical thinking, project complex text onto a potential topic space to obtain the "document-topic" distribution θ and the "topic-feature word" distribution ϕ [31], as shown in [Figure 4: see original paper]. The number of topics can be determined based on the perplexity evaluation method [32]:

$$\text{perplexity}(D) = \exp\left(-\frac{\sum_{d=1}^{M} \log p(w_d)}{\sum_{d=1}^{M} N_d}\right) \quad (2)$$

$$p(w) = p(z|d) \cdot p(w|z) \quad (3)$$

where M is the number of texts, D is the test set document, N_d is the total number of all words appearing in document d, and $p(w)$ is the probability of each word appearing in the test set. $p(z|d)$ is the probability of each topic appearing in the document, and $p(w|z)$ represents the probability of each feature word appearing under a certain topic [32]. Perplexity decreases with the increase of the number of topics K, and the K value when the decline trend becomes flat is the optimal number. Then, sort the words under topics according to relevance, take the top N as feature words, and form N "topic-feature word" distributions.

Step 3: After obtaining the "document-topic-feature word" matrix, if the sample size is small, each topic word cluster can be manually identified, considering both the distribution and semantic relationships of words under the topic. Since social academic Apps are mostly vertical industry platforms (e.g., DXY App for the medical field and Jingguanzhijia App for the economics and management field), domain experts can be consulted to identify topic feature words and set domain labels that can summarize the characteristics for each document. If the sample size is large, classification algorithms in machine learning such as KNN (k-Nearest Neighbor) can be used for document label matching.

Step 4: After obtaining the domain labels of each document, digitize them by assigning values. For example, assign values $\{1, 2, 3\}$ to the three labels "Internal Medicine," "Surgery," and "Traditional Chinese Medicine," ultimately achieving the goal of identifying user interest fields from user-generated content and converting them into numerical values.

The above method can quantify text data and summarize text features from the topic level, balancing practicality and scientificity. It achieves the research goal of applying text data together with other data to portrait depiction and can maximize the avoidance of portrait distortion.

---

4 Key Technology Introduction and Framework Model Formation

4.1 Key Technologies for Dynamic Portrait Construction

Using clustering algorithms to divide large-scale users into several typical groups can efficiently grasp user core characteristics and needs in data-driven environments [22]. This idea has been widely applied: for example, Chen Tianyuan used K-means clustering to depict differentiated user groups in mobile libraries [33]; Chen Juan et al. used hierarchical clustering to identify three typical groups on the Zhihu platform [34].

Social academic App users have large user bases, making them suitable for mining group portrait features through clustering methods. However, unlike previous clustering for static data, the dynamic portrait proposed in this study introduces time segments to depict user dynamic trajectories by identifying migration relationships between clusters in adjacent time periods. Currently, few studies have explored dynamic portrait methods based on time series data, but scholars widely agree that clustering time series data needs to consider two core issues: Clustering results within a specific time period should fully reflect the characteristics of data within that period; Clustering results in different time periods should show certain continuity on the time axis, i.e., clusters in adjacent time periods evolve smoothly [35]. D. Chakrabarti et al. first proposed the evolutionary clustering idea in 2006 and applied it to the K-means algorithm to obtain evolutionary K-means clustering to solve the accuracy and continuity problems of time series data clustering [36]. On this basis, Wang Fupeng considered the impact of historical data on current clustering results and applied it to trend analysis of financial stock market trajectories to help investors understand stock market changes in real time [37].

Since the research objectives and data structure of this study are similar to the above research, this study believes that evolutionary clustering ideas can be fully borrowed to dynamically mine the evolutionary behavior of social academic App users. The steps are as follows:

Step 1: Divide reasonable time windows. Divide the collected data into multiple groups according to a certain time period t, or collect data multiple times according to time periods to obtain multiple groups of data in different time periods. The t value can be determined based on the App product development or iteration cycle, or can be determined by reference to the loss function [35], i.e.:

$$\text{Cost} = \alpha \cdot C_S + (1 - \alpha) \cdot C_T$$

where C_S is the snapshot cost (cost of snapshot), with larger values indicating worse clustering effects in the current time period; C_T is the temporal cost (cost of temporality), with larger temporal costs indicating worse smoothness between adjacent time periods. α is the weight coefficient balancing these two values. The larger α is, the more attention is paid to clustering quality within time periods; the smaller α is, the more attention is paid to temporal smoothness.

Step 2: Construct user time-specific portraits. This process borrows the evolutionary K-means algorithm proposed in literature [37] to achieve, i.e., incorporating consideration of the weights of cluster centers in historical time periods into the selection of cluster center points in traditional K-means algorithm clustering. The formula is:

$$C_t^j \leftarrow \alpha \times E_{x\in \text{closest}(j)}(x) + (1 - \alpha) \times \sum_{i=1}^{t} (f(t, t-i) \cdot C_{t-i}^j) \quad (4)$$

where C_t^j represents the jth cluster center point in time period t, E is the expectation operation, closest(j) represents sample points closest to center point

j, and $f(t, t-i)$ refers to the weight of the cluster center in time period $t-i$. α remains the weight coefficient balancing snapshot loss and temporal loss. In addition, when selecting initial center points, there is no historical data for reference, and the K-means algorithm has the drawbacks of requiring manually specified cluster numbers and randomly selected initial center points. Therefore, the Canopy algorithm can be used as a prior cluster basis before K-means clustering [38], while determining points with high Silhouette coefficients and reasonable sample distribution in each group as initial center points [39], and then completing the clustering of time series data.

Step 3: Identify evolutionary trajectories. This step requires analyzing the mapping relationships between clusters in adjacent time periods, i.e., judging the similarity of clusters, which can be achieved by calculating cluster weights through conditional probability. The specific method is: given clustering results Q_i and Q_{i+1} at two adjacent moments t_i and t_{i+1}, the weight calculation formula for cluster $C_m(t_i)$ in Q_i and cluster $C_u(t_{i+1})$ in Q_{i+1} [37] is:

$$\text{Weight}(C_m(t_i), C_u(t_{i+1})) = P(X \in C_u(t_{i+1}) | X \in C_m(t_i)) = \frac{\sum P(x \in C_m(t_i) \cap C_u(t_{i+1}))}{\sum P(x \in C_m(t_i))} \quad (5)$$

where $P(X \in C_u(t_{i+1}) | X \in C_m(t_i))$ represents the probability of sample points belonging to $C_u(t_{i+1})$ under the condition of belonging to $C_m(t_i)$, $P(x \in C_m(t_i) \cap C_u(t_{i+1}))$ is the probability of sample points belonging to $C_m(t_i) \cap C_u(t_{i+1})$, and $P(x \in C_m(t_i))$ is the probability of sample points belonging to $C_m(t_i)$. Due to data imbalance within time periods, clusters generally have seven evolutionary states, as shown in [Figure 5: see original paper]: emergence of a new group; a group splitting into two or more groups; two or more groups merging into one group; a group disappearing in the next stage; the number of users in a group increasing in the next stage; the number of users in a group decreasing in the next stage; and a group remaining unchanged in adjacent stages. Before identifying evolutionary behavior, critical thresholds for cluster evolution need to be set. The MEC framework (Monitor of the Evolution of Clusters over time) proposed by M. OLIVEIRA et al. identifies various evolutionary behaviors by defining survival parameter τ and split parameter ϕ [40]. Wang Fupeng introduced parameter μ as the critical value for grow and decline behaviors to identify the above seven evolutionary behaviors. The thresholds for all three parameters are [0, 1], and specific values should be determined through repeated iterative experiments according to actual application scenarios [37].

Social academic App users' behavioral habits often have continuity. Evolutionary clustering considers the influence of historical data on current clustering results, making it more suitable for depicting the actual state of social academic App users. Moreover, the evolutionary K-means algorithm incorporating the Canopy algorithm has stronger robustness [38]. In addition, as users' interaction with the platform deepens, users' belonging groups are in a dynamic change process, which is consistent with cluster evolution analysis. Therefore, using the evolutionary K-means algorithm to implement social academic App dynamic portraits is theoretically highly feasible, and relevant applied research [37] provides practical verification for it.

4.2 Framework Model of Social Academic App User Dynamic Portrait Based on Small Data

Due to different construction objectives and application scenarios, the hierarchical structure and specific construction methods of portrait models also differ. Y. Kritikou et al. believe that models should include three layers: monitoring layer, modeling layer, and adaptation layer [41]; Xu Pengcheng et al. constructed a digital library user portrait framework model from six layers: data collection, processing, storage, mining, presentation, and application [25]. Drawing on the above research, this study believes that the dynamic portrait of social academic App users based on small data essentially collects small data strongly related to users based on the tag system, constructs user time-specific portraits by introducing time windows, and identifies user dynamic trajectories through cluster migration. Therefore, the framework model should include three layers: small data collection and processing layer, time-specific portrait construction layer, and dynamic portrait formation layer, as shown in [Figure 6: see original paper]:

4.2.1 Small Data Collection and Processing Layer The small data collection and processing layer obtains relevant small data through comprehensive use of technologies such as web crawlers and log mining, as well as surveys that can obtain users' deep characteristics, and performs preprocessing to transform it into data forms that meet portrait needs. Since the correlation between each dimension tag and portrait results varies, and the degree of correlation is related to the specific application scenario of the dynamic portrait, tag weights should be adjusted according to actual needs to ensure the scientific decision-making value of portrait results.

4.2.2 Time-Specific Portrait Construction Layer Constructing user time-specific portraits is the prerequisite for identifying behavior trajectories. First, data is divided into specific time periods according to time information to form time series data. Then, clustering is performed on data within different time periods respectively, i.e., users with similar characteristics within a specific time period are divided into the same cluster, and cluster centers are determined. Finally, the time-specific portrait results are automatically stored in the database to pave the way for the next step of identifying user dynamic behavior trajectories.

4.2.3 Dynamic Portrait Formation Layer Forming dynamic portraits requires completion through three parts: similarity judgment, threshold comparison, and dynamic trajectory identification. First, calculate the weights of clusters in adjacent time periods, then compare the obtained weights with preset thresholds to judge migration relationships such as birth, merging, splitting, and death between clusters, and identify user dynamic migration trajectories. To ensure the accuracy and practicality of portraits, portrait results are evaluated. If they meet platform decision-making requirements, portrait results are stored for utilization; if not, feedback is used to adjust the organizational content and process structure at each level to construct reliable and applicable dynamic portraits.

The above three layers form a closed-loop process that not only conforms to the general process of portrait models but also achieves the research goal of depicting social academic App user dynamic portraits based on small data. Moreover, the portrait evaluation module considers the practical application value of portraits, making it more scientific.

4.3 Application Value of Social Academic App User Dynamic Portrait Based on Small Data

4.3.1 Predicting User Behavior Conducting dynamic portraits of existing platform users can predict behavioral characteristics in the next stage by observing their dynamic trends, effectively overcoming the lag shortcomings of traditional portrait methods. At the same time, by comparing features or calculating center point distances, new users can be positioned to the most similar cluster in the time-specific portrait database, and then user behavior in the next time period can be predicted according to the trajectory model, which can solve the cold start problem for new users to a certain extent.

4.3.2 Optimizing Precision Services The social academic App user dynamic portrait model constructed in this study uses small data as portrait data support, considering both users' surface behavioral characteristics and deep behavioral driving factors. The depicted portraits are more aligned with users' actual states and comprehensive demands. Social academic App operators can thus refine user groups and recommend adaptive content according to group characteristics, achieving precise matching between users and resources, thereby optimizing platform precision service levels and improving decision-making efficiency and capability.

4.3.3 Adjusting Product Layout The identification and tracking of user behavior trajectories enable App platforms to grasp user needs in real time and dynamically. Constructing dynamic portraits helps customize products and services covering the entire user-platform interaction process for users, and can also assist platforms in adjusting product layout proactively according to the evolution of user needs, optimizing platform structure, dynamically connecting with users' personalized needs, and continuously improving platform market competitiveness.

Conclusion

As a continuation and supplement of big data, small data has strong superiority in comprehensively and deeply representing individual user behavior patterns and contextual factors. Meanwhile, the maturity and popularization of wireless sensor technology, smart wearable devices, and monitoring and positioning technologies provide technical support for real-time acquisition of small data. In view of this, this study proposes the idea of depicting social academic App user portraits based on small data, explores the feasibility of using evolutionary clustering and cluster migration to achieve dynamic portraits, constructs a framework model for social academic App user dynamic portraits based on small data, and elaborates on its role in behavior prediction, service optimization, and layout adjustment. It provides new perspectives and reference ideas for social academic App platform operators to accurately understand users' dynamic needs and timely formulate adaptive resource service strategies.

Since current research on constructing social academic App user dynamic portraits from the small data perspective is very scarce, this study aims to propose a comprehensive and systematic framework and conduct feasibility analysis to enrich and expand the theoretical research system of small data dynamic portraits, providing new perspectives for the application of small data and innovative breakthroughs in user portraits. Due to space limitations, the relevant empirical process cannot be fully presented in one article. Follow-up research will conduct empirical analysis of small data collection and processing and dynamic portrait construction based on this study to deeply explore the application value and generalization effect of the model.

References

[1] Geng Bin, Sun Jianjun. Research on user behavior of online academic social platforms: Taking Nanjing University users on the ResearchGate platform as an example[J]. Library and Information, 2017, 61(5): 47-53.

[2] HOLDEN R J, KULANTHAIVEL A, PURKAYASTHA S, et al. Know thy Health user: development of biopsychosocial personas from a study of older adults with heart failure[J]. International journal of medical informatics, 2017, 108(12): 158-167.

[3] TRUSOV M, MA L, JAMAL Z. Crumbs of the cookie: user profiling in customer-base analysis and behavioral targeting[J]. Marketing science, 2016, 35(3): 405-426.

[4] Chen Chen, Ma Xiaoting. Research on precise user portrait of library based on small data[J]. Information and Documentation Services, 2018(5): 57-61.

[5] Sun Danxia, Wang Weijun, Jiang Yi. Research on embedded subject service based on user small data[J]. Library Work and Research, 2019(4): 84-90.

[6] Diao Yu, Chang Pei. Research on precision maker services in libraries oriented to small data[J]. Library Theory and Practice, 2018(5): 109-112.

[7] Wang Xin, Zhang Dongmei. Research on personalized intelligent service of library based on small data of university readers under big data environment[J]. Information Studies: Theory & Application, 2018, 41(2): 132-137.

[8] ESTRIN D. Small data, where n=me[J]. Communications of the ACM, 2014, 57(4): 32-34.

[9] Chen Chen, Ma Xiaoting. Research on precise user portrait of library based on small data[J]. Information and Documentation Services, 2018(5): 57-61.

[10] HSIEH C K, YANG L, WEI H, et al. Immersive recommendation: news and event recommendation using personal digital traces[C]//Proceedings of the 25th international conference on World Wide Web. Montreal: ACM, 2016: 51-62.

[11] Chen Chen, Li Qiang. Reader interest discovery and prediction based on small data decision-making[J]. Information Science, 2017, 35(5): 75-80.

[12] Li Lirui, Deng Zhonghua. Research on small data fusion of scientific research users under the "Internet+" background[J]. Library and Information Service, 2016, 60(6): 58-63.

[13] Zhang Xiaodan, Jiang Hong, Wang Kehui. Empirical study on influencing factors of academic App user adoption intention[J]. Library and Information Service, 2018, 62(18): 90-101.

[14] Ji Qingnan. Research on intelligent bus APP interaction design based on user experience[D]. Xi'an: Xi'an University of Technology, 2017.

[15] Li Dawei, Du Hongbo, Zhou Xiaolin, et al. Design of book recommendation App based on "user portrait" mining[J]. Software, 2018, 39(5): 35-37.

[16] Han Zhangjunjie. Design and application of user portrait for information App based on data analysis[D]. Beijing: University of Chinese Academy of Sciences, 2017.

[17] Guo Wei. Research on development method of sci-tech journal APP based on cloud platform: Taking "Changbaishan Academic Hub" as an example[J]. Chinese Journal of Scientific and Technical Periodicals, 2018, 29(5): 485-490.

[18] Wang Hui. Research on influencing factors of academic journal APP usage[J]. Journal of Southwest Petroleum University (Social Sciences Edition), 2017, 19(6): 76-82.

[19] Liu Yong, Wu Xiangyu, Xie Benju. Research on information recommendation based on dynamic user portrait[J]. Computer Systems & Applications, 2018, 27(6): 236-239.

[20] Wang Yicheng, Wang Ping. Research on construction of sci-tech intelligence service recommendation model based on user dynamic portrait[J]. Information Studies: Theory & Application, 2019, 42(4): 83-88.

[21] Zhang Huimin. Research on dynamic user portrait based on lifestyle transformation[D]. Wuxi: Jiangnan University, 2018.

[22] Zhang Liman, Zhang Xiangxian, Lu Heng, et al. Research on paid user group portrait of knowledge live broadcast platform[J]. Library and Information Service, 2019, 63(5): 84-91.

[23] Liu Man. Empirical study on determinants of college students' information search behavior based on TPB[J]. Library Work and Research, 2014(5): 39-44.

[24] LEWIN K. Field theory in social science[M]. New York: Harper and Brother Publishers, 1951: 239-240.

[25] Xu Pengcheng, Bi Qiang, Zhang Han, et al. Construction of digital library user portrait model under data-driven environment[J]. Library and Information Service, 2019, 63(3): 30-37.

[26] Heng Shupeng, Zhou Zongkui, Lei Yuju, et al. The effect of actual-ideal self-discrepancy on adolescent game addiction: The serial mediating effect of avatar identification and immersion[J]. Psychological Research, 2018, 16(2): 253-260.

[27] Shi Xiaojiao. Research on the influence mechanism of management team cognitive ability on decision-making effect[D]. Shenyang: Shenyang University of Technology, 2018.

[28] Zhang Jidong, Luo Shasha. Research on mobile library user behavior perception based on contextual preference[J]. Information Science, 2018, 36(9): 52-56.

[29] Du Wei, Gao Changyuan. Research on personalized context recommendation model in mobile e-commerce environment[J]. Information Studies: Theory & Application, 2017, 40(10): 56-61.

[30] Yu Chuanming, Tian Xin, Guo Yajing, et al. Research on user portrait based on behavior-content fusion model[J]. Library and Information Service, 2018, 62(13): 54-63.

[31] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 3(1): 993-1022.

[32] Zeng Ziming, Wan Pinyu. Public security event microblog sentiment analysis integrating evolutionary features[J]. Information Science, 2018, 36(12): 3-8, 51.

[33] Chen Tianyuan. Empirical study on user portrait construction of university mobile library[J]. Library and Information Service, 2018, 62(7): 38-46.

[34] Chen Juan, Wu Zhuoqing, Deng Shengli. User segmentation and behavior analysis of "Zhihu" based on hierarchical clustering method[J]. Information

Studies: Theory & Application, 2018, 41(7): 111-116.

[35] Lin Huihui. Research on evolutionary clustering algorithm and its application[D]. Yangzhou: Yangzhou University, 2017.

[36] CHAKRABARTI D, KUMAR R, TOMKINS A. Evolutionary clustering[C]//Proceedings of the 12th ACM SIGKDD conference on Knowledge discovery and data mining. New York: ACM, 2006: 554-560.

[37] Wang Fupeng. Research on evolutionary clustering and its application in financial stock market[D]. Hangzhou: Zhejiang University, 2014.

[38] Zhang Lin, Mou Xiangwei. Chinese text clustering algorithm based on Canopy+K-means[J]. Library Tribune, 2018, 38(6): 113-119.

[39] Lu Jianyun, Zhu Qingsheng, Wu Quanwang. A heuristic method for determining the number of clusters[J]. Small Microcomputer System, 2018, 39(7): 1381-1385.

[40] OLIVEIRA M, GAMA J. A framework to monitor clusters evolution applied to economy and finance problems[J]. Intelligent data analysis, 2012, 16(1): 93-111.

[41] KRITIKOU Y, DEMESTICHAS P, ADAMOPOULOU E, et al. User profile modeling in the context of Web-based learning management systems[J]. Journal of network & computer applications, 2008, 31(4): 603-627.

Author Contributions

Zhang Xiangxian: Guidance and determination of the paper framework;
Zhang Liman: Proposal of research ideas and initial draft writing;
Wu Yawei: Paper revision and proofreading;
Guo Shunli: Paper revision and proofreading.

Abstract (English)

Purpose/Significance: To enrich and expand the theoretical research system of building dynamic portraits based on small data, and provide ideas and references for social academic App platforms to effectively predict the evolution trend of user behavior and improve precision service levels. **Method/Process:** Firstly, based on an in-depth analysis of the concept and characteristics of small data, combined with the features of social academic Apps, a dynamic portrait label system is designed from two aspects of user surface behavior and deep driving factors. Then, small data with strong correlation and high value with users is collected as the data support of the portrait, and the acquisition and processing methods of portrait small data are clarified. Finally, the research method for implementing dynamic portraits is proposed and an overall framework model is formed. **Result/Conclusion:** The construction of dynamic portraits of social academic App users based on small data can effectively refine portrait granularity and improve the lag of previous portraits, which has important reference value for promoting the accurate service level of social academic App platforms under data-driven situations.

Keywords: small data; social academic App; dynamic portrait; behavior prediction

Note: Figure translations are in progress. See original paper for figures.